MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2019/2020

DSP5014 – STATISTICS AND PROBABILITY

(for diploma students only)

2 MARCH 2020
9.00 a.m. – 11.00 a.m.
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of 5 pages excluding cover page and appendix.
2. Attempt **ALL** questions. The distribution of the marks for each question is given.
3. Please write all your answer in the Answer Booklet provided.
4. Key formulae are given in the Appendix.

QUESTION 1

a) Provide an example for each of the following data type: (3 marks)

i) Qualitative.

ii) Quantitative Discrete.

iii) Quantitative Continuous.

b) Diagram 1 shows a pie chart on Cheryl's monthly expenditure. Her total monthly expenditure is RM1300.

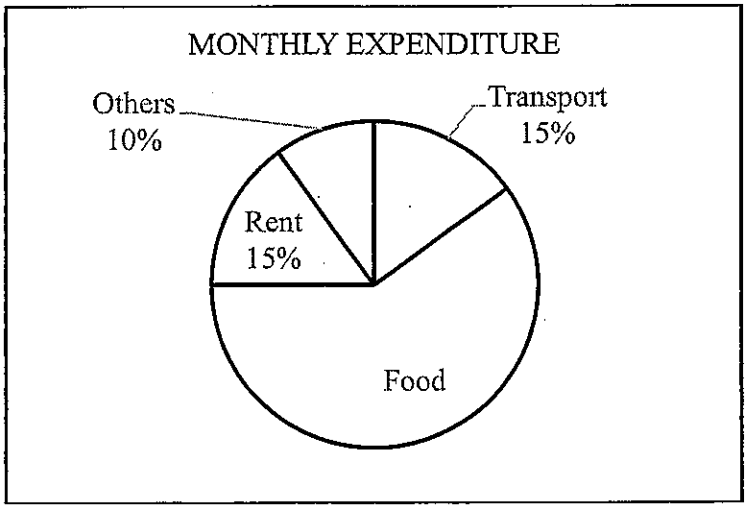

Diagram 1

i) Find the percentage of food expenditure. (1 mark)

ii) Construct a distribution table consisting type of expenditure, frequency and relative frequency to represent the data above. (4 marks)

c) Given a data set consisting number of people living in a house: 8, 14, 7, 6, 5. Find the

i) median. (2 marks)

ii) mean. (2 marks)

iii) shape of distribution. (2 marks)

Continued ...

d) Given a data set of time taken by 30 students in a Race Run 400m.

46	75	48	52	67	51	75	55	32	54
58	33	65	73	59	70	56	71	55	73
34	62	76	78	68	79	39	72	64	46

i) Construct a grouped data table starting with $30 \leqslant 40$. (4 marks)

Based on your table in (i),

ii) state the mode class. (1 mark)

iii) calculate the variance. (4 marks)

e) Diagram 2 shows a stem-and-leaf of marks for 14 students who failed in their midterm test.

Stem	Leaves
1	5
2	2 2 4 8
3	2 5 9
4	1 3 5 7 7 9

Key: 1 | 5 = 15

Diagram 2

i) Find the quartiles. (6 marks)

ii) List out the 5-number summary. (1 mark)

[TOTAL: 30 MARKS]

QUESTION 2

a) List out **THREE** properties of probability. (3 marks)

b) A box contains 2 pieces of black cards, 9 pieces of pink cards and x pieces of white cards. When a card is chosen at random from the box, the probability of getting a pink card is $\frac{3}{4}$. Find the value of x. (4 marks)

Continued ...

c) During Teacher's day celebration in SMK Bukit, each class are required to send a representative for a singing competition. The competition is divided into 2 categories which are Lower Form (L) category and Upper Form (U) category. Each category consists of female (F) and male (M) students. Draw a tree diagram to represent the event. (3 marks)

d) Table 1 shows a random sample of 100 hikers and their hiking area preferences.

Gender	Hill (H)	Lake (L)	Beach (B)	Total
Male (M)	18	16	x	45
Female (F)	16	y	14	55
Total	34	41	25	100

Table 1

Find the value of

i) x and y. (2 marks)

ii) $P(F \cup L)$. (2 marks)

iii) $P(B \mid M)$. (3 marks)

iv) $P(H \cap B)$ if both events are independent. (3 marks)

[TOTAL: 20 MARKS]

QUESTION 3

a) A normal tomato plant will have a mean diameter of 8.2cm and a standard deviation of 2.4cm. A biotechnologist claimed that more larger tomatoes can be produced with the use of nanotechnology. To check on this, a random sample of 36 tomatoes was taken and gave a sample mean of 9.1cm. At 5% significance value, can we accept the claim by the biotechnologist? (10 marks)

Continued ...

b) A previous survey done by Student Representative Council (SRC) regarding "the effectiveness of clubs and societies in making student understand event management better" showed that less than 48% students agree with this statement. To gain more information on the current situation, SRC has conducted another survey and the result is shown as below:

YES	YES	NO	NO	YES	NO	YES	YES	NO	NO
NO	NO	YES	YES	NO	YES	NO	NO	NO	YES
YES	NO	NO	NO	NO	NO	NO	YES	YES	NO
NO	NO	YES	NO	NO	YES	YES	YES	NO	YES

Based on the result above, can we conclude that the previous study is true at 10% significance level? (12 marks)

c) A car magazine is comparing the average expenses per day for two models of sports car due to the popular opinion that Model X has lower expense per day compared to Model Y. A random sample of Model X and Model Y gives a test statistics value of $z_0 = -2.16$ and p-value 0.0154. At 5% significance level, can we conclude that Model X has lower expenses than Model Y? (3 marks)

[TOTAL: 25 MARKS]

QUESTION 4

A final year student in Diploma of Fashion and Design would like to study the relationship between shoe size and height. He recorded the details of 10 persons as follows.

Shoe size, x	Height (cm), y
6.5	166
9	168
8.5	166.5
10.5	170
9.5	170
7.5	167.5
12	171
9.5	169
4.5	164.5
5	166

a) Calculate SS_{xx}, SS_{yy} and SS_{xy} based on the information given. (9 marks)

b) Calculate the correlation coefficient, r and interpret the answer. (4 marks)

Continued ...

c) Find the linear regression equation $\hat{y} = \beta_0 + \beta_1 x$. (5 marks)

d) Predict the height of a person if the shoe size is 8. (2 marks)

e) Compute the value of *SST*, *SSR* and *SSE*. (3 marks)

f) Calculate the standard error of the estimate, s_e. (2 marks)

[TOTAL: 25 MARKS]

End of Paper

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$ where n : sample size	$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$ where m : class midpoint f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{\left(\sum f_i\right)-1}\left[\sum m_i^2 f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}\right]$ where m : class midpoint f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \frac{x}{n}$
2	$Z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$Z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\hat{p}_p\left(1-\hat{p}_p\right)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ *where* $\hat{p}_p = \frac{x_1 + x_2}{n_1 + n_2}$

Continued...

LINEAR REGRESSION ANALYSIS

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \dfrac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$

- **Correlation coefficient,** $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$	$S_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n}$	$S_{xy} = \sum xy - \dfrac{(\sum x)(\sum y)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \dfrac{(S_{xy})^2}{S_{xx}}$

- **Test of Significance for Regression Slope**

Standard Error, $s_e = \sqrt{\dfrac{SSE}{n-2}}$	Regression t-Test $t = \dfrac{\beta_1}{s_e / \sqrt{Sxx}}$	Correlation t-Test $t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$

End of Paper